Knowledge graph-based recommendation framework identifies drivers of resistance in EGFR mutant non-small cell lung cancer

Dr Dimitris Polychronopoulos (Oncology R&D)

Dr Anna Gogleva (R&D IT)

17th July 2022, ICML, Baltimore, USA

Global, science-led, patient-focused biopharmaceutical company

Science and innovation-led

Therapy areas of focus:
Oncology;
Cardiovascular, Renal & Metabolism;
Respiratory & Immunology;
Rare Disease

Diversified portfolio with broad coverage across primary care, specialty care and rare diseases

Commitment to people and society

Global strength, with balanced presence across regions

Strategic R&D sites close to global bioscience clusters

New Haven

Gothenburg

Boston
New York

South San Francisco

Shanghai

Osaka

Cambridge

Gaithersburg

Focus on main therapy areas and key platforms

Oncology

Cardiovascular, Renal & Metabolism

Respiratory & Immunology

Rare Disease

Combination of capabilities

Small molecules	Biologics	Protein engineering	Complement inhibition	Other emerging drug platforms	Diagnostics	Devices

Oncology

We are leading a revolution in oncology to redefine cancer care and Data Science plays a critical part

Our clinical strategy is designed to help transform survival

With our portfolio and pipeline we strive to revolutionise cancer care

Catalysing changes in the practice of medicine to transform the patient experience

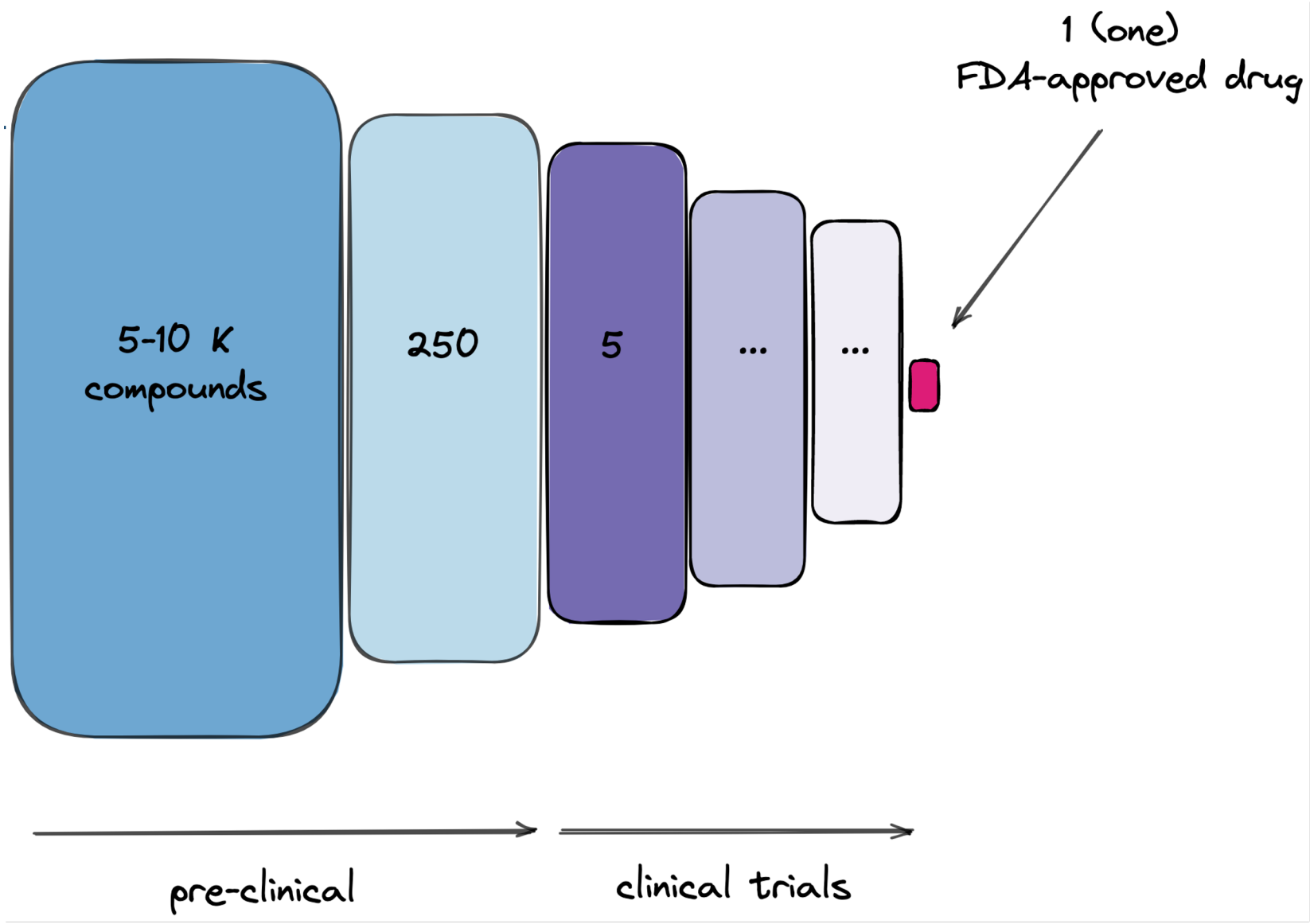
We are driven by our passion, our people and a culture of innovation

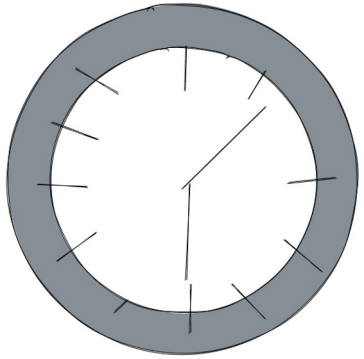
Life-cycle of a medicine

entire life-cycle of a medicine:

- research and development
- manufacturing and supply
- global commercialisation

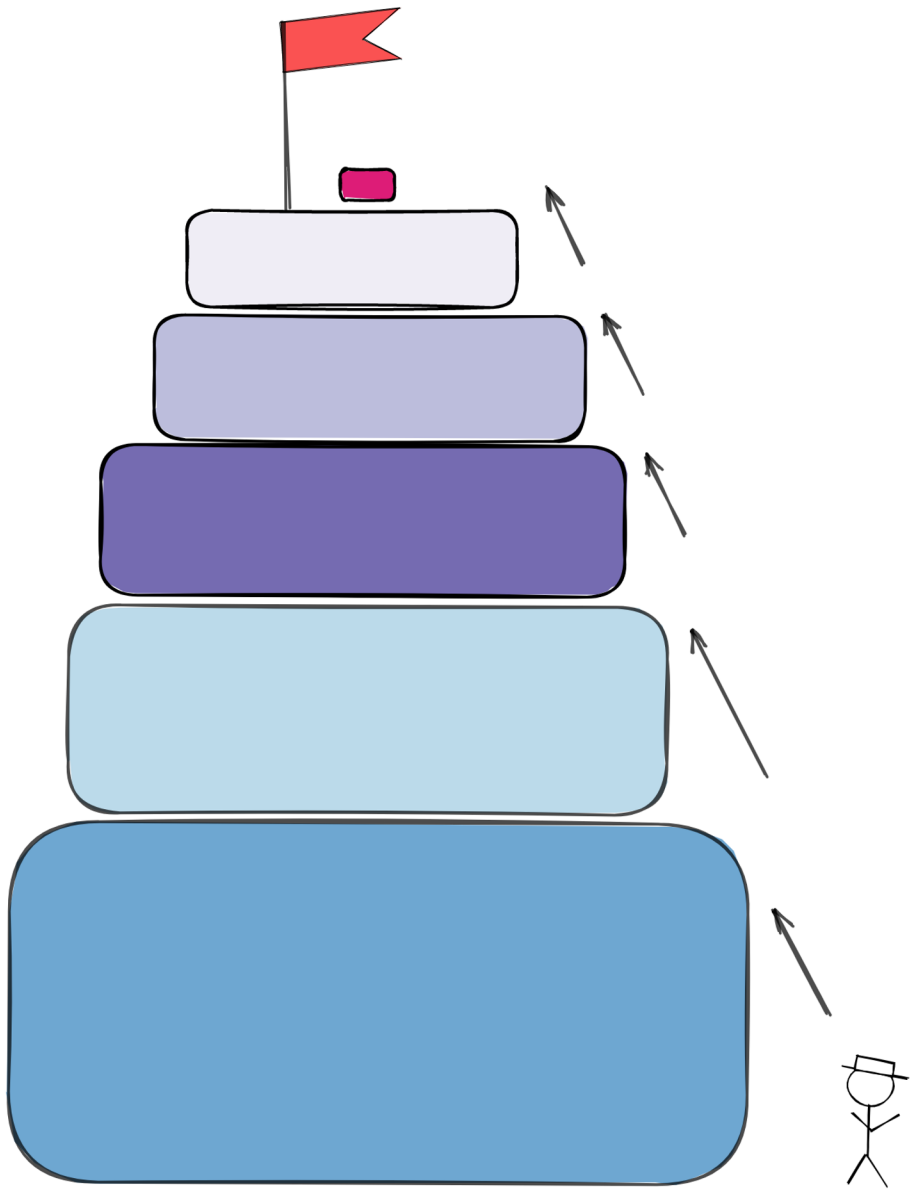
One needs to fail a lot to discover a working drug





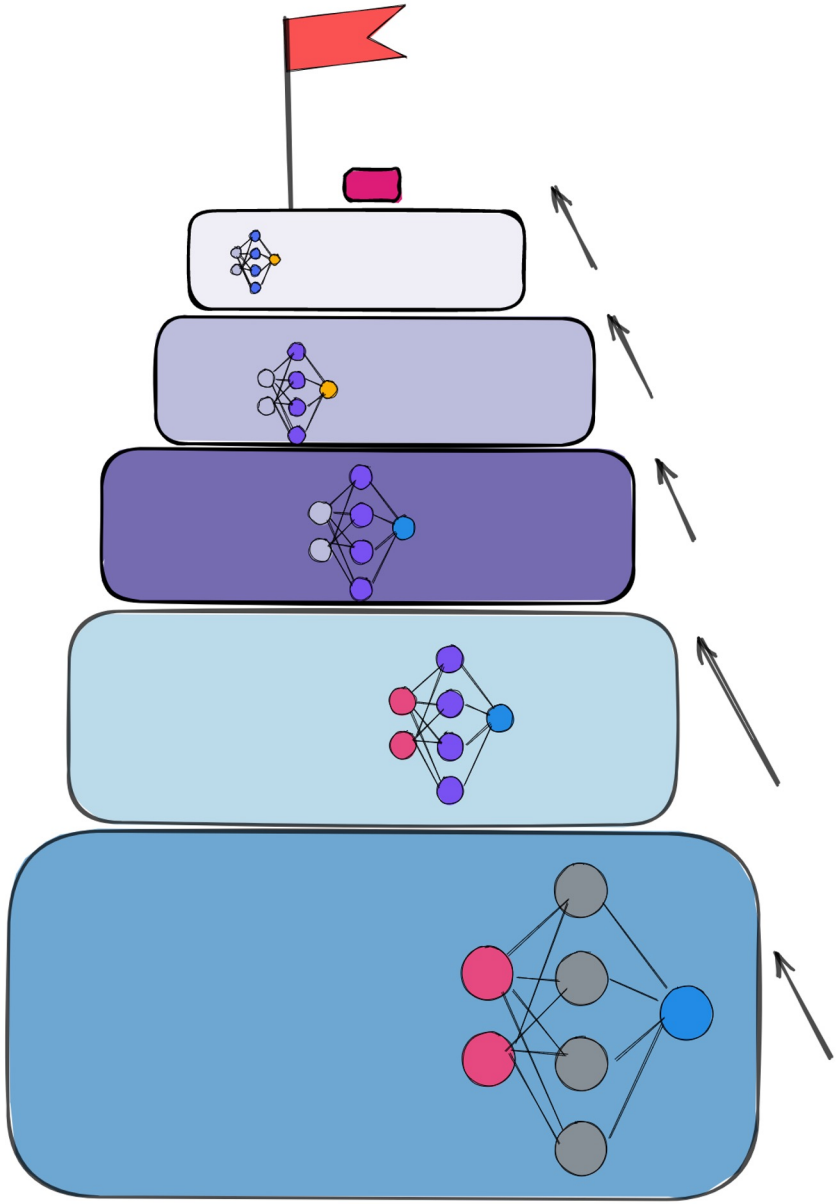
6 - 10
years

It is a tall mountain to climb

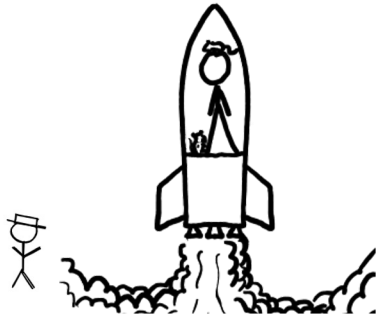


- How to develop new efficient treatments faster?
- How to make better decisions in the process?

It is a tall mountain to climb

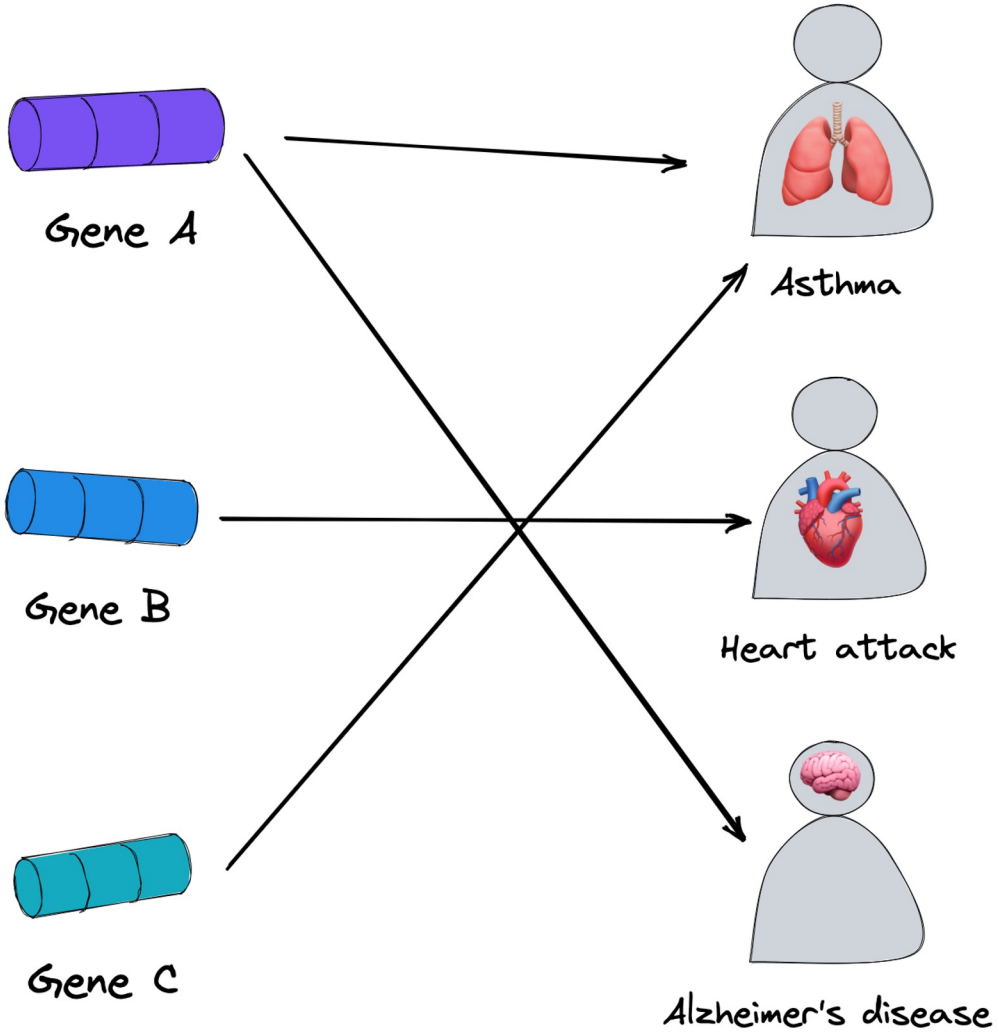


- How to develop new efficient treatments faster?
- How to make better decisions in the process?
- Recommendation systems can help in multiple places

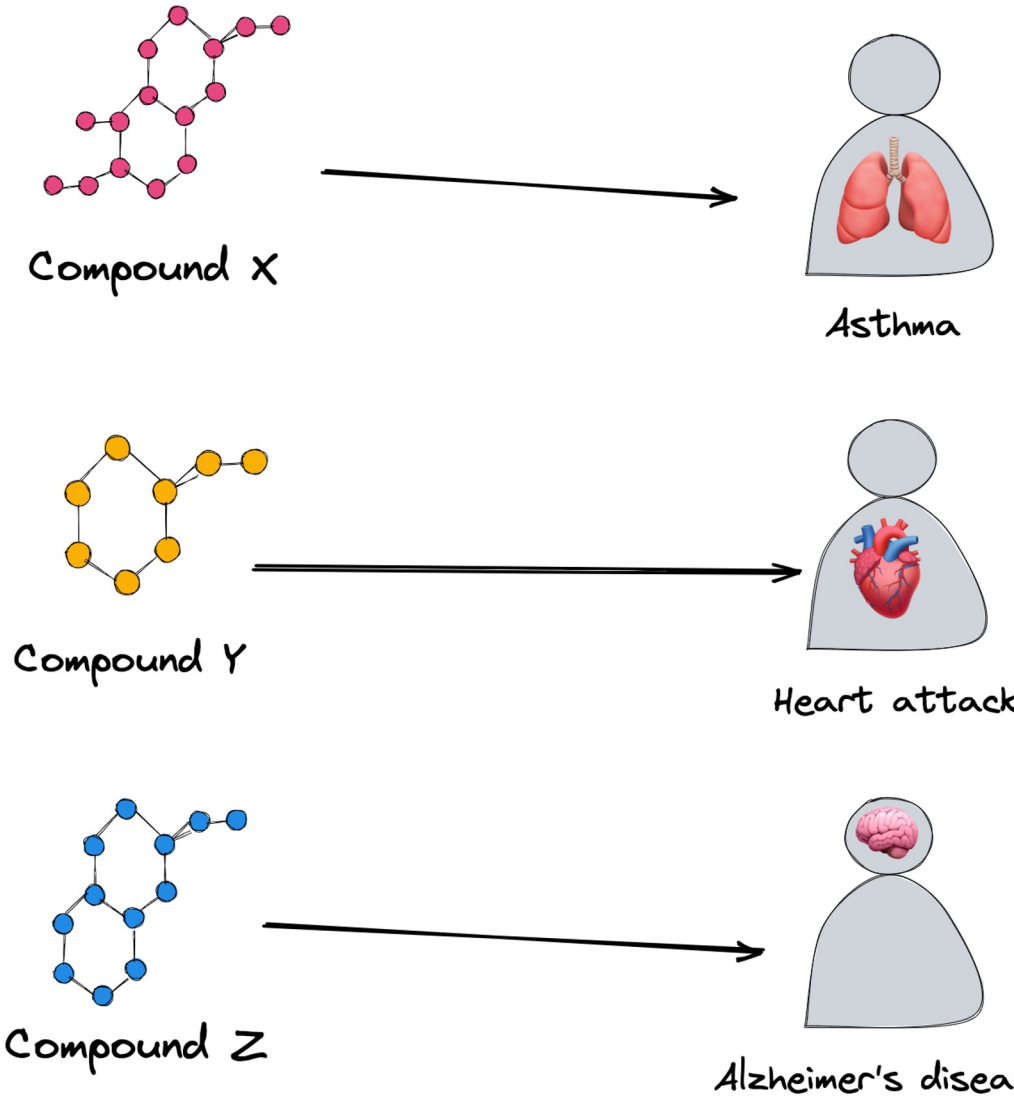


Recommendation problems in drug discovery

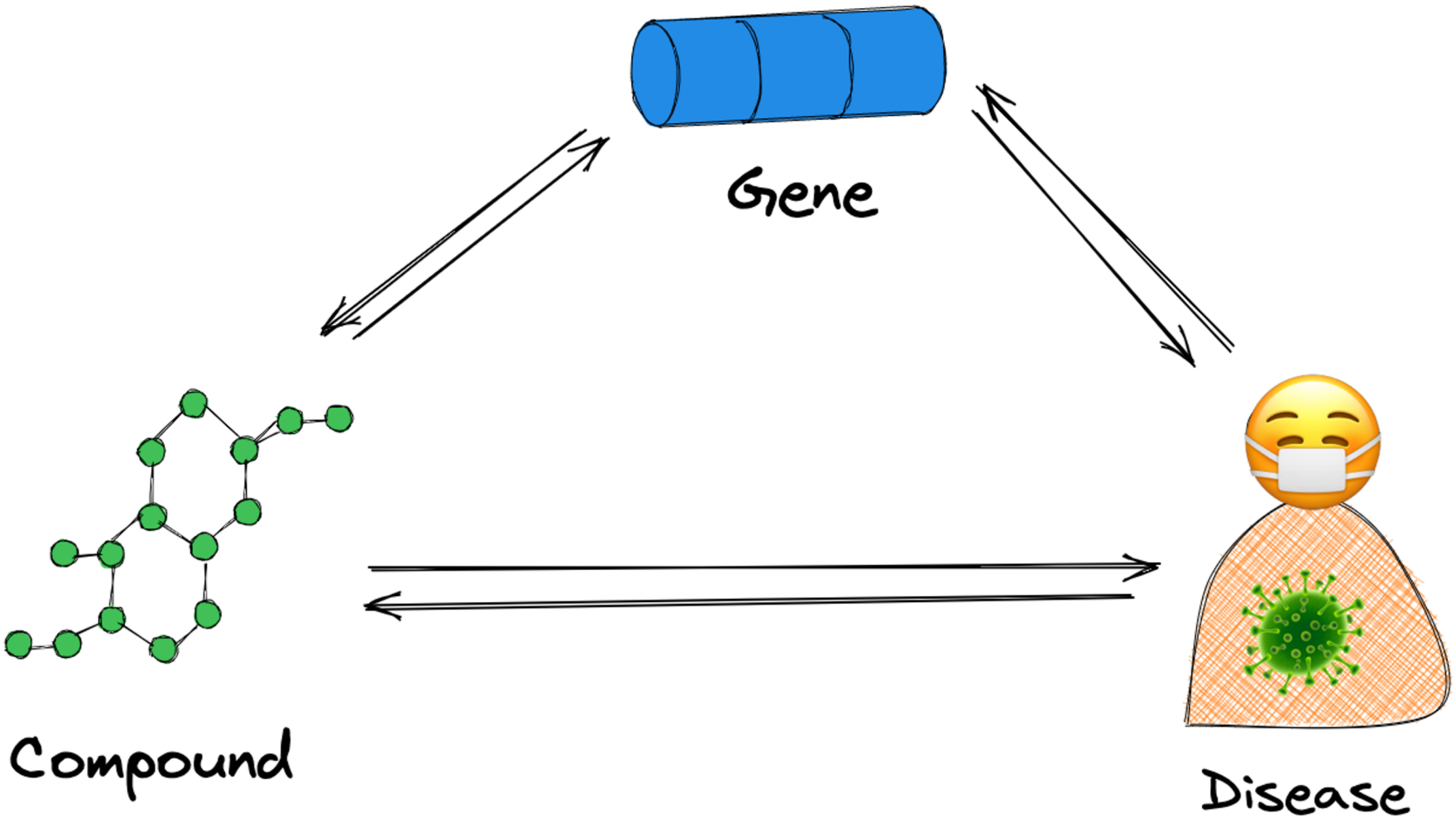
find a gene causing a disease



match a drug with a disease



Drugs, genes, diseases

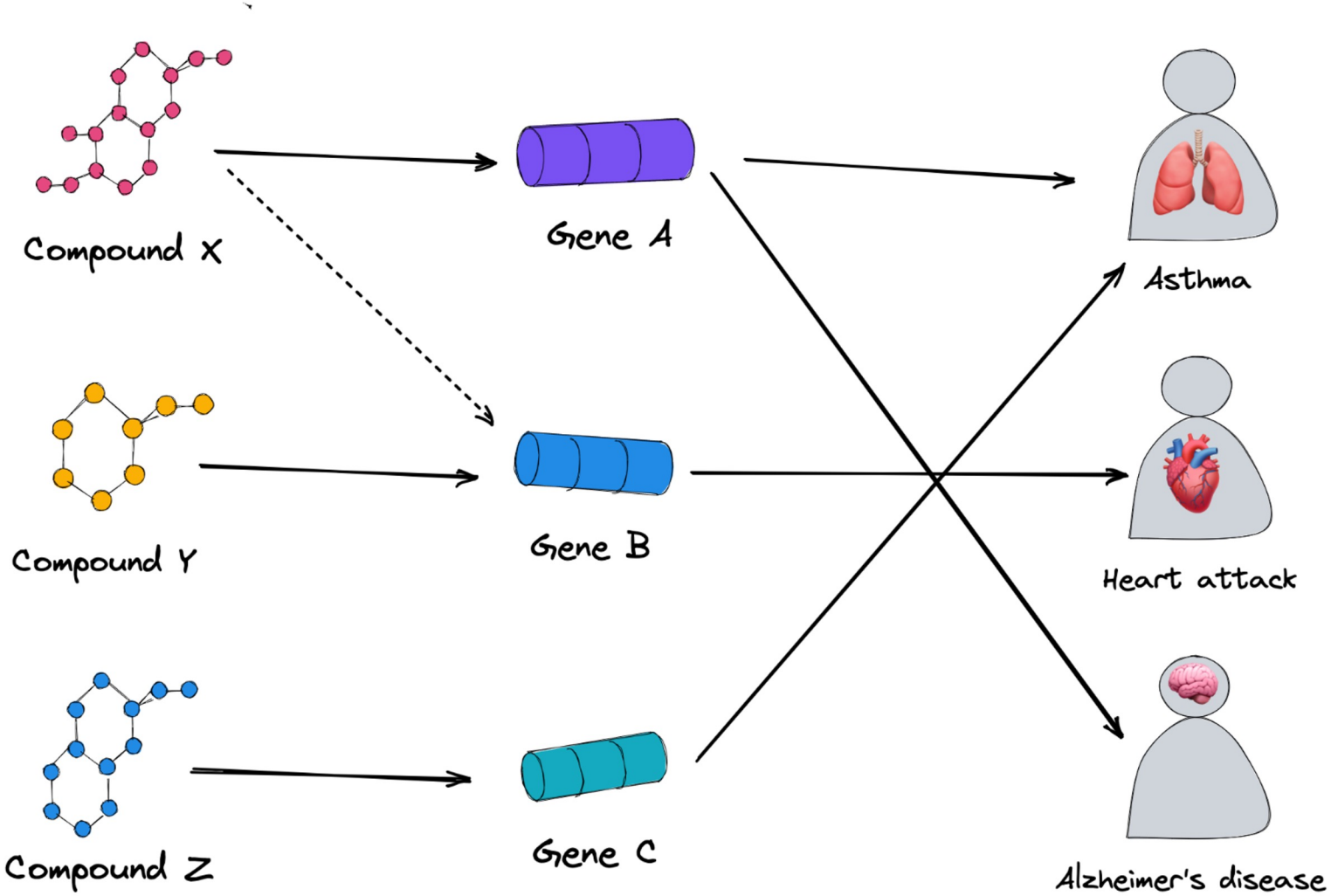


It gets complex very fast

Millions of compounds
Billions possible theoretically

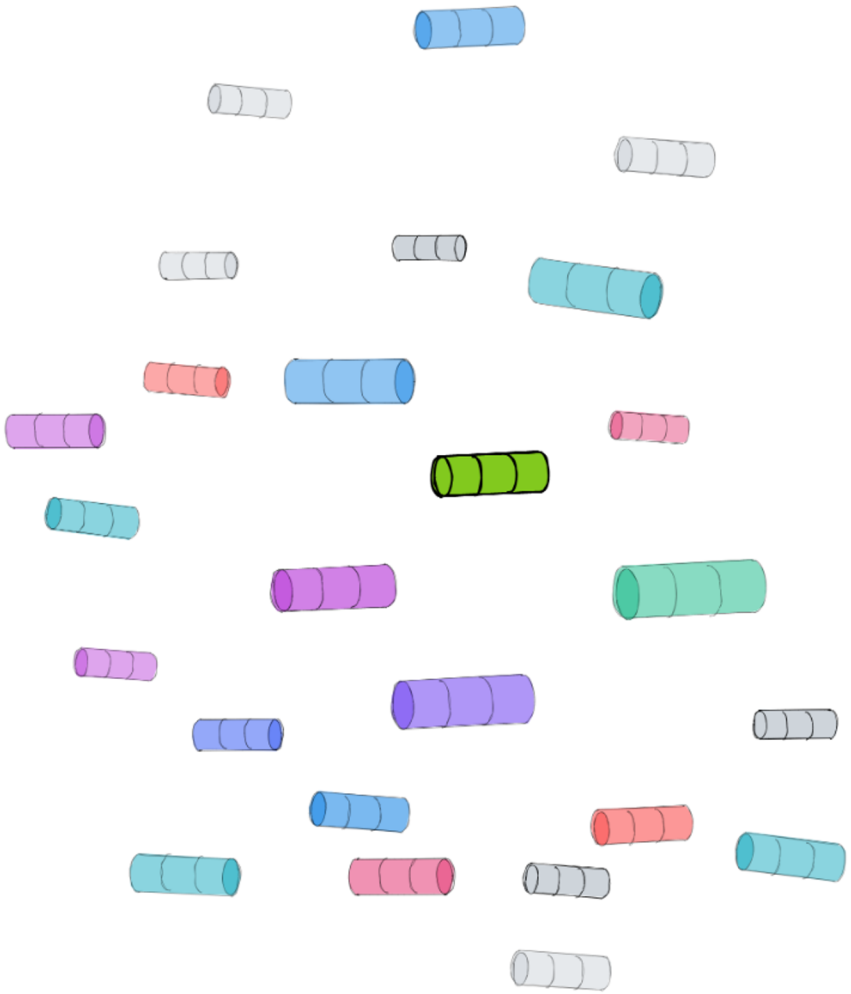
25-30 K genes,
80 K functional elements

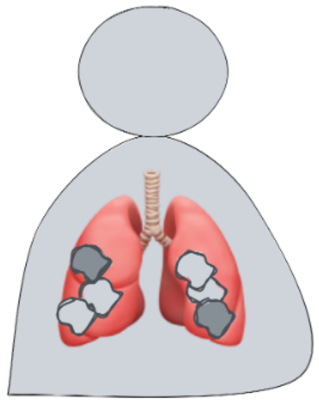
~10 K diseases



It is rarely just a single gene

- 25-30K human genes
- everything interacts with everything,
  each gene is a suspect

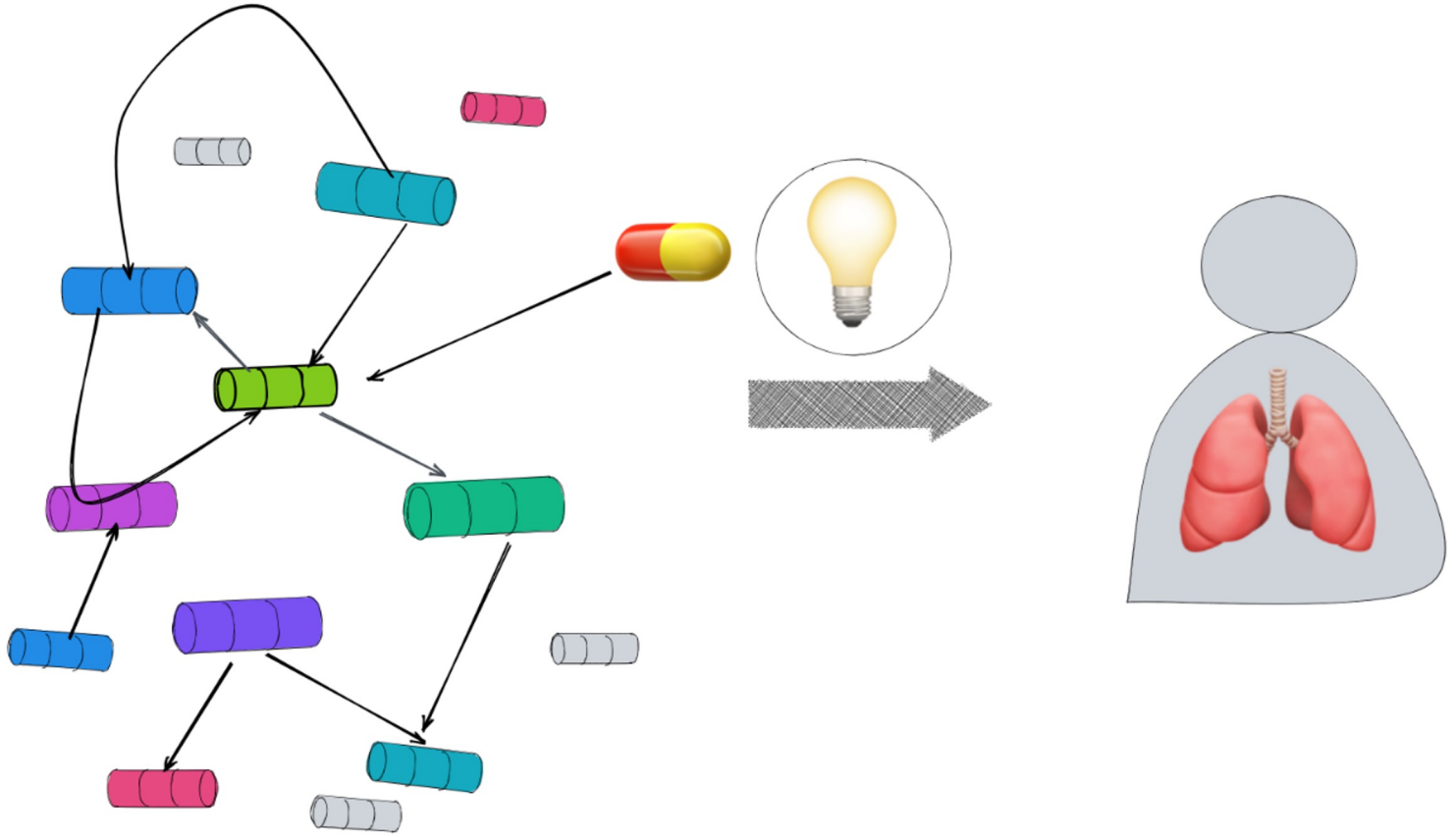




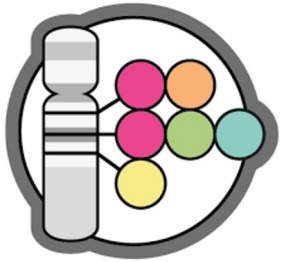
a disease

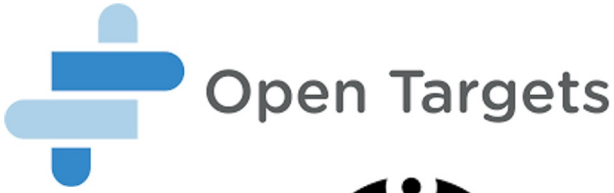
Find a molecular network behind a disease

1. disease ~ a molecular process gone awry
2. find the key molecular process
3. re-route it safely

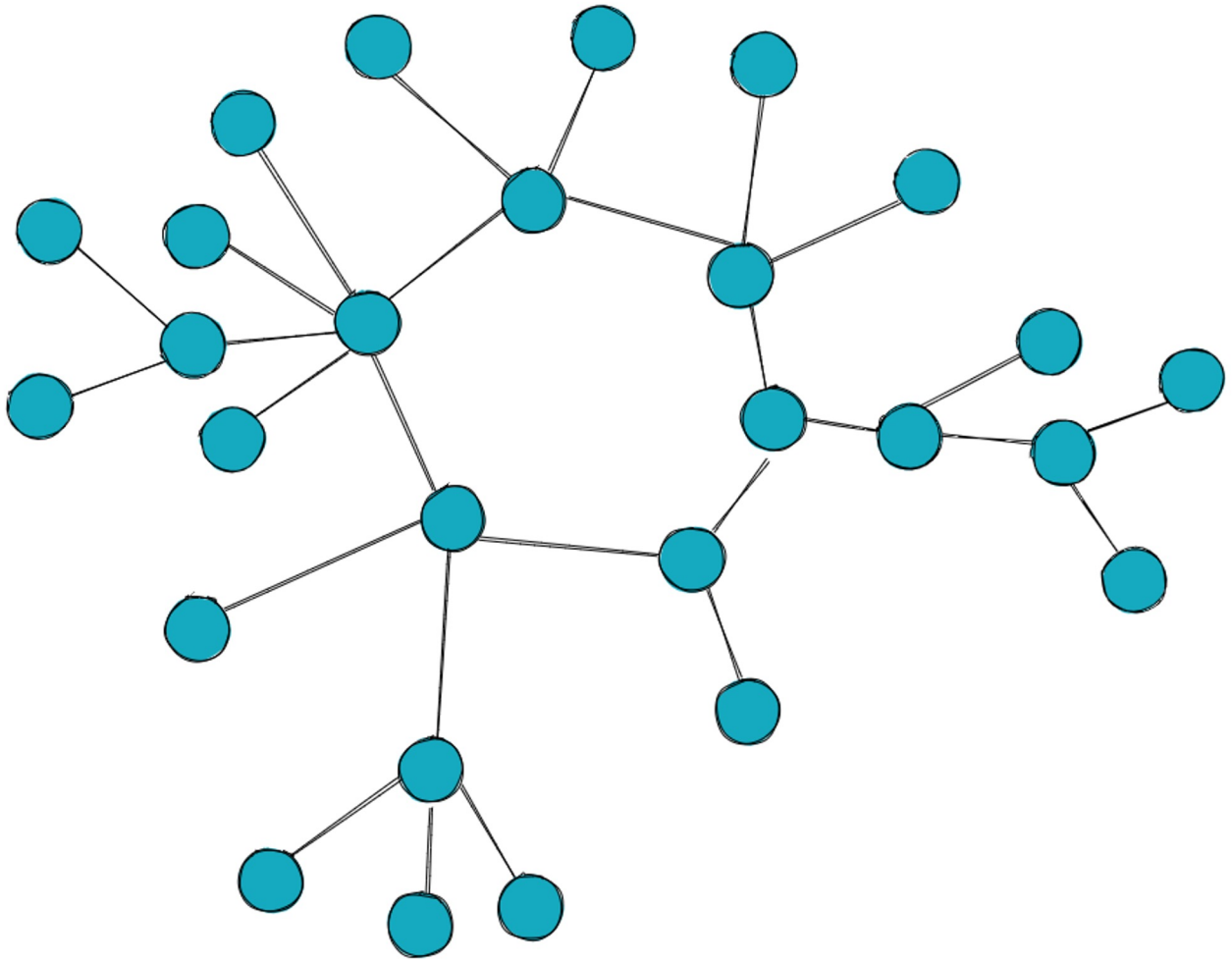


Biomedical knowledge is spread across multiple resources





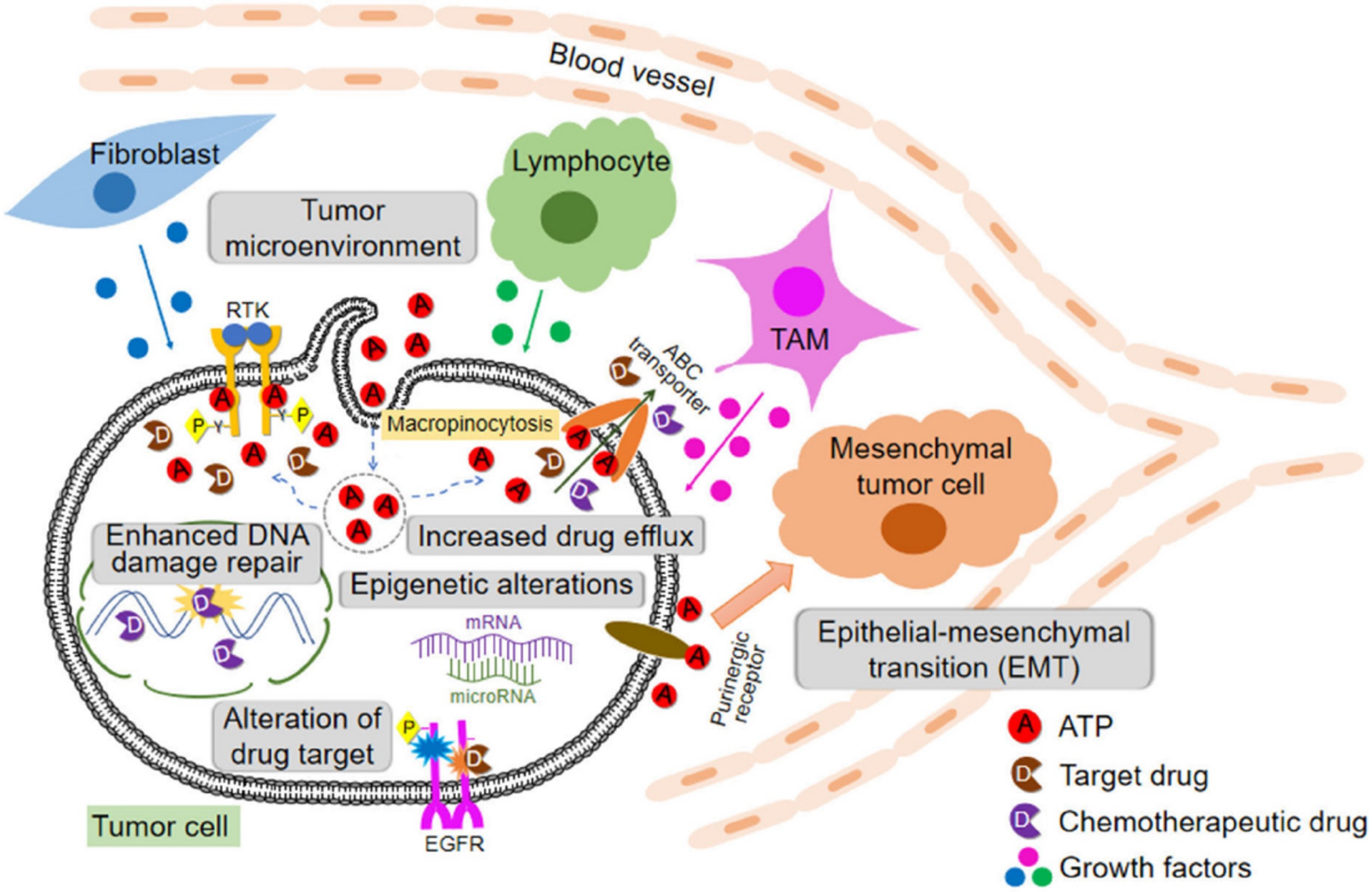
Graph makes things simpler



- Biomedical information often comes in forms of **networks** and **hierarchies**
- Graph is a convenient way to organize it
- BIKG (our internal knowledge graph): **60+** data sources including - omics and data extracted from the literature
- **11 M nodes, 1 B edges**
- Use graph as a source of context and features for recommenders

Early success story: graph-based recommendations

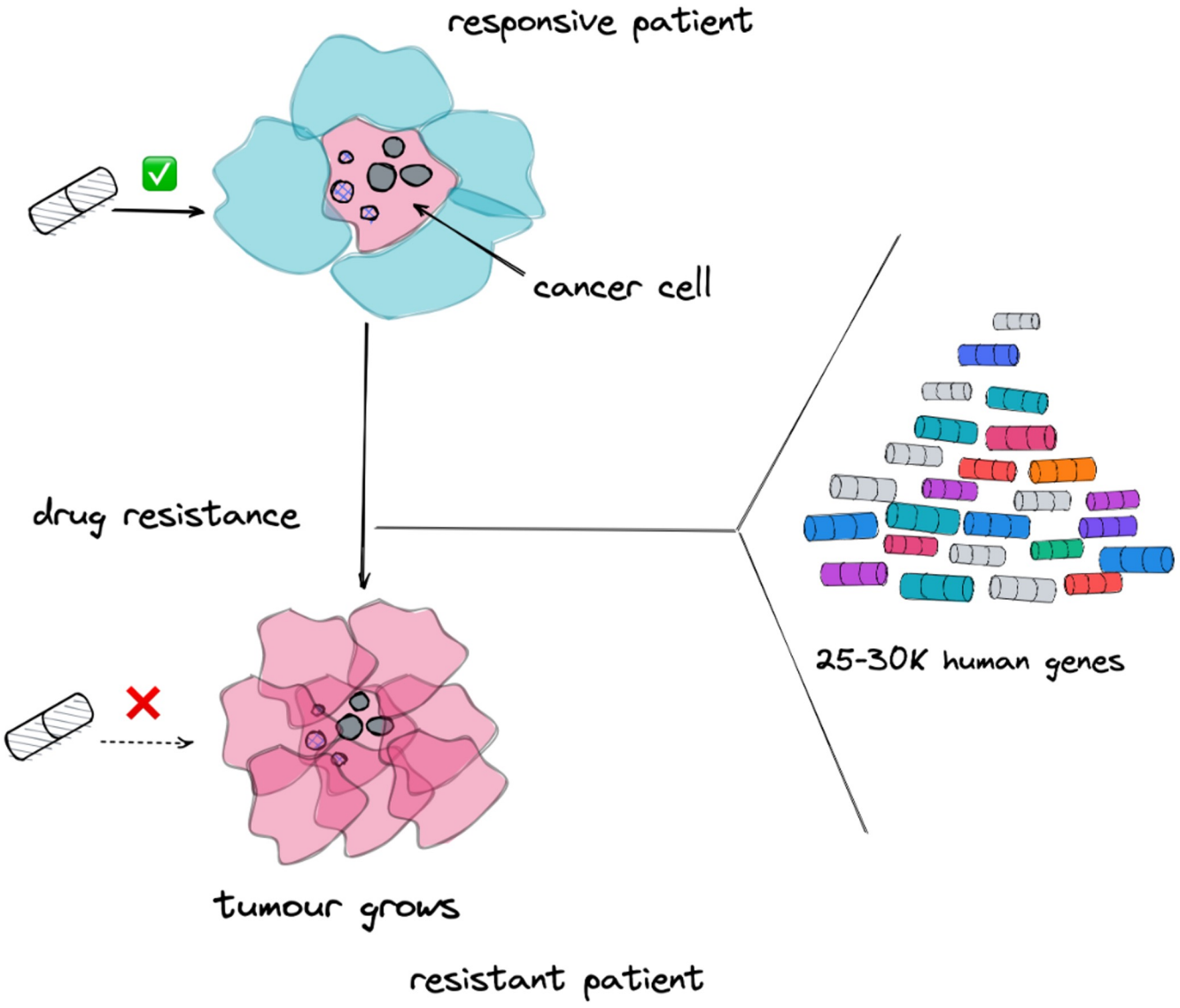
Applied recommendation problem: contextualize experimental data



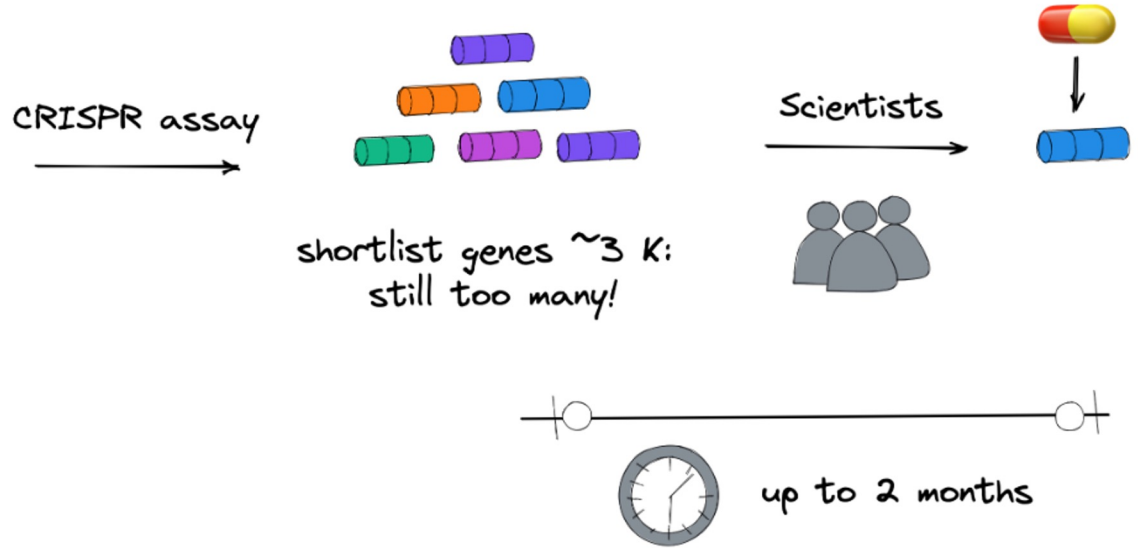
- Drug resistance in lung cancer
- Occurs in a sub-population of patients
- Resistance landscape is complex

X Wang, H Zhang, X Chen - Cancer Drug Resistance, 2019

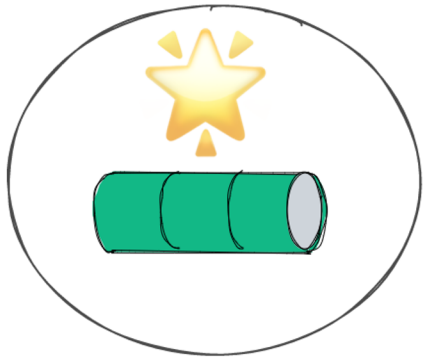
How to help scientist find key genes faster?



- key gene == target
- find a target gene ➡️ develop new treatments to overcome the resistance



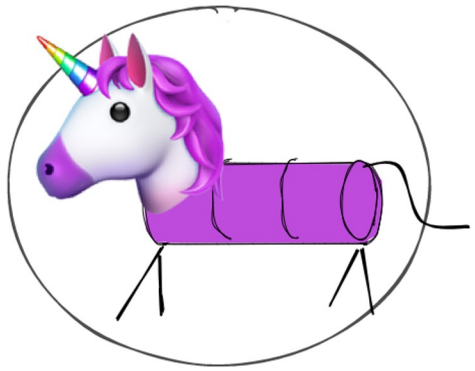
An ideal target



- [x] Expression
- [x] Pathway/complex enrichment
- [x] Effect size
- [x] Druggability
- [x] Mode of action
- [x] Translation in models
- [x] Internal assets
- [x] Bench validation
- [x] Consistency in assays
- [x] Clinical relevance
- [x] Literature support
- [x] Novelty

...

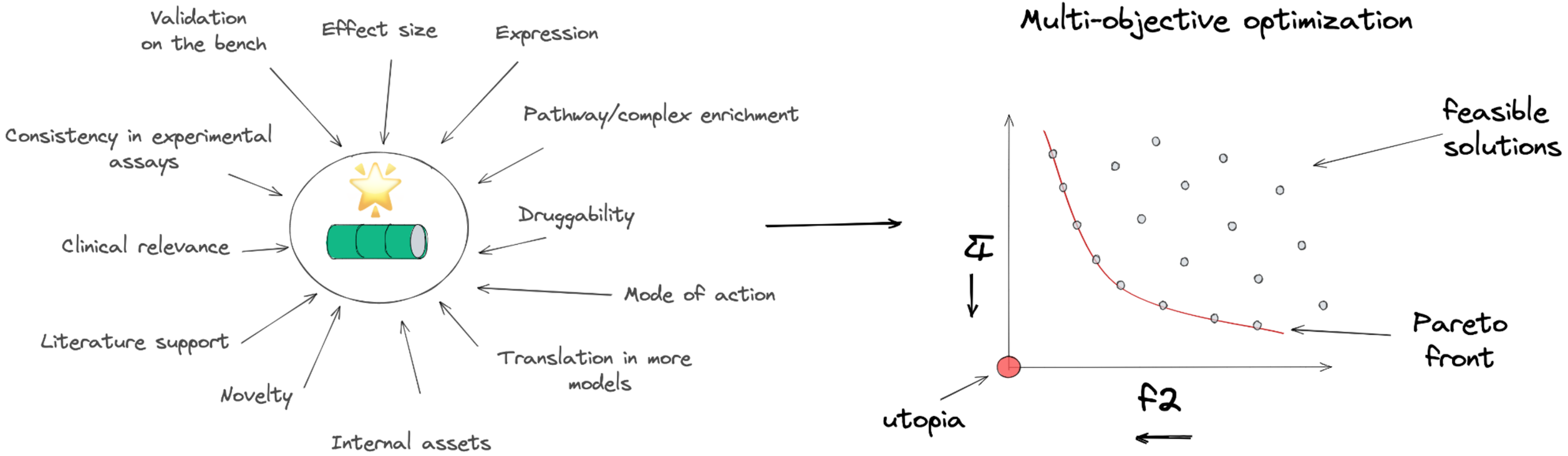
An ideal target does not exist



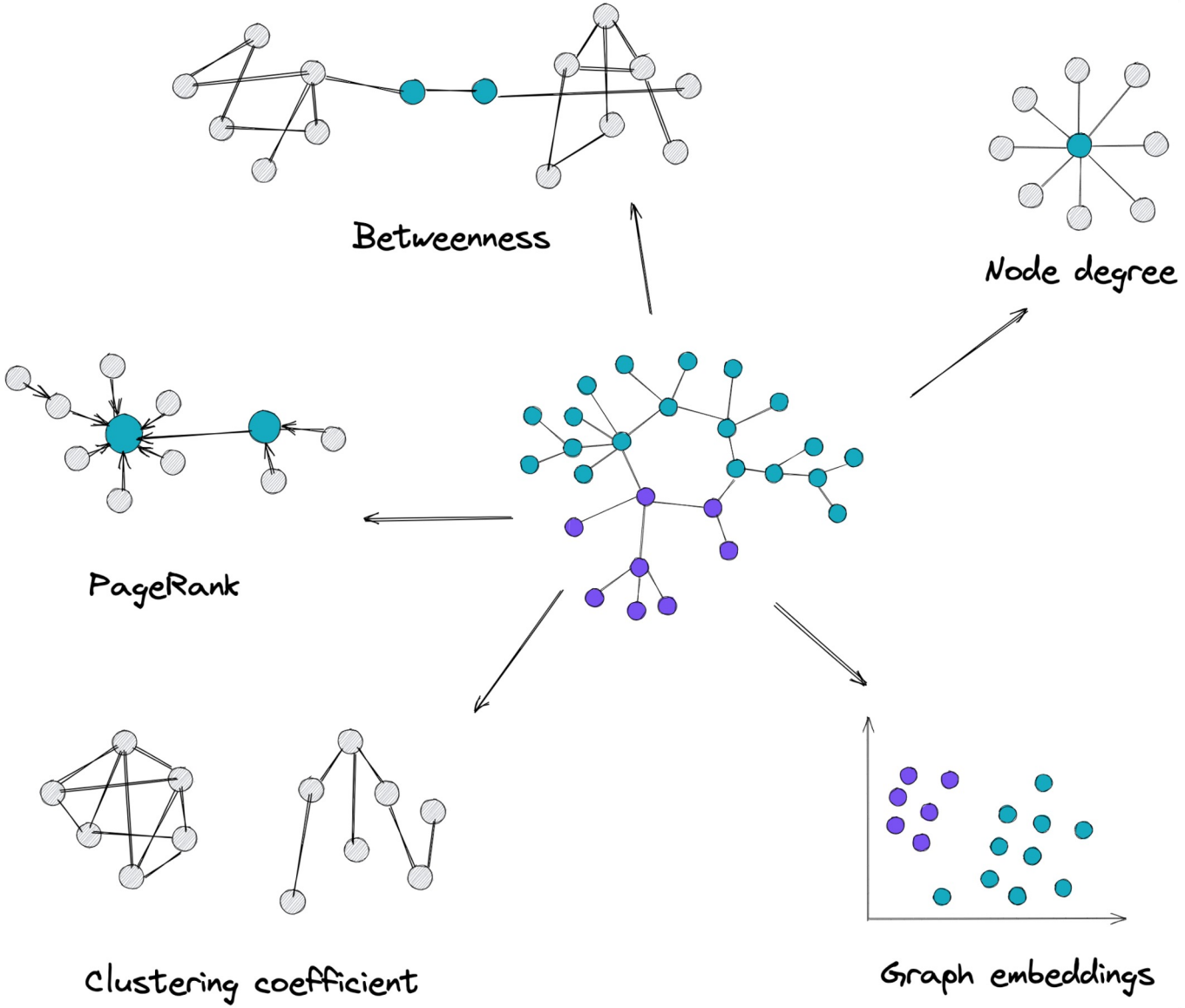
- [x] Expression
- [x] Pathway/complex enrichment
- [x] Effect size
- [x] Druggability
- [x] Mode of action
- [x] Translation in models
- [x] Internal assets
- [x] Bench validation
- [x] Consistency in assays
- [x] Clinical relevance
- [x] Literature support
- [x] Novelty

...

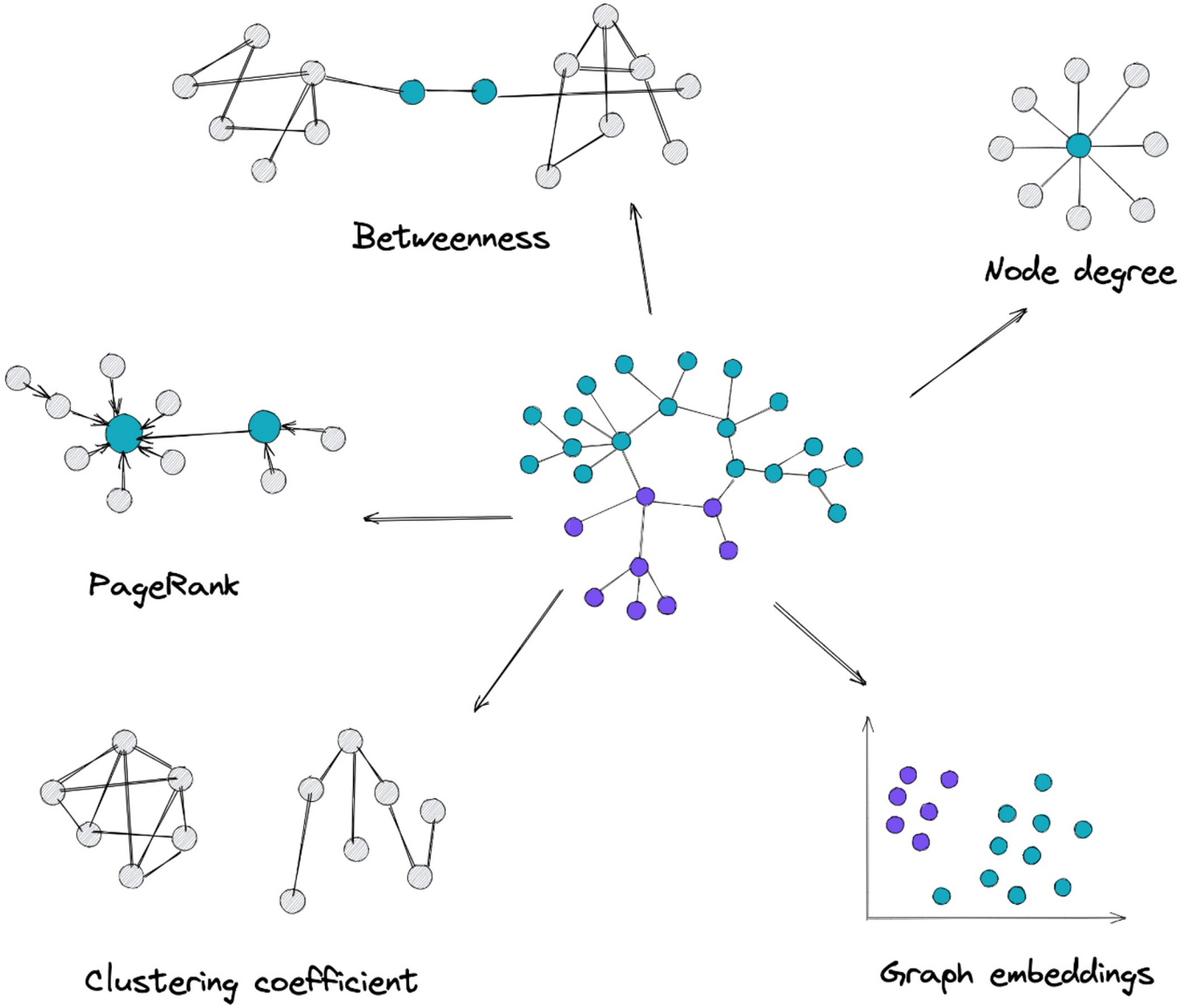
Target selection as an optimization problem



Hybrid feature set: source features from the graph

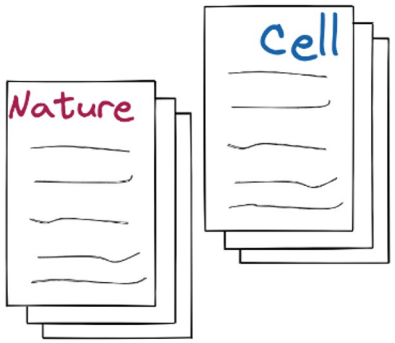


Hybrid feature set: combine with clinical features



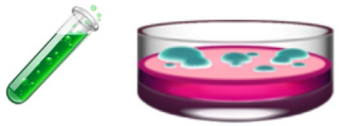
Clinical features

Literature support



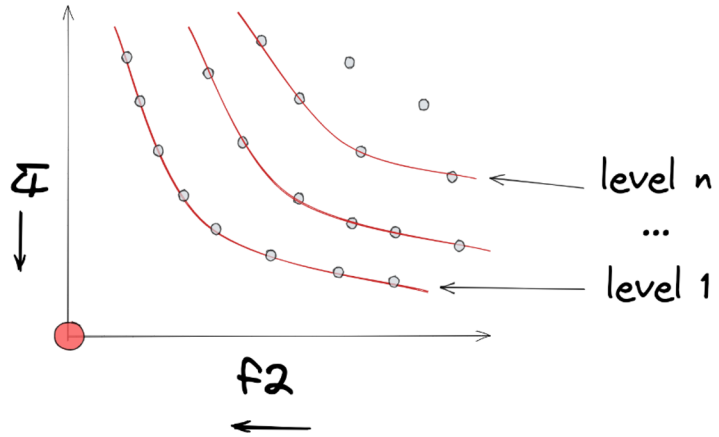
Druggability

Pre-clinical experimental assays

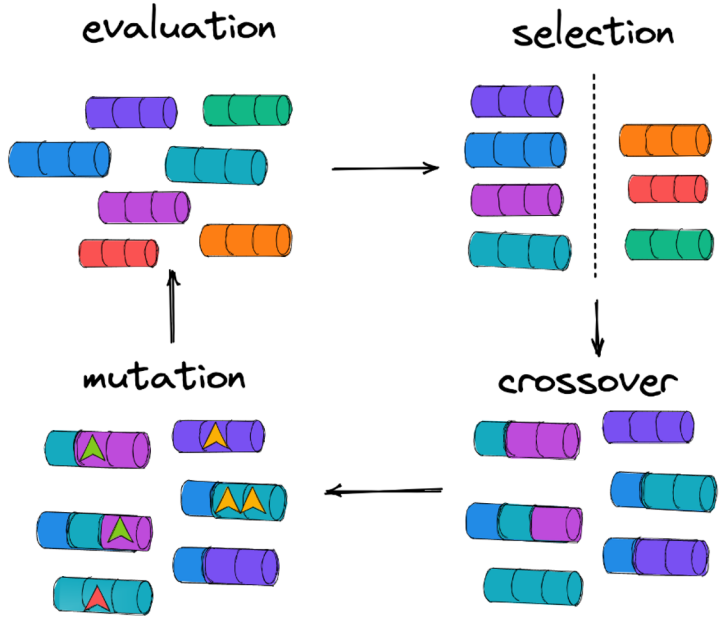


Approaches

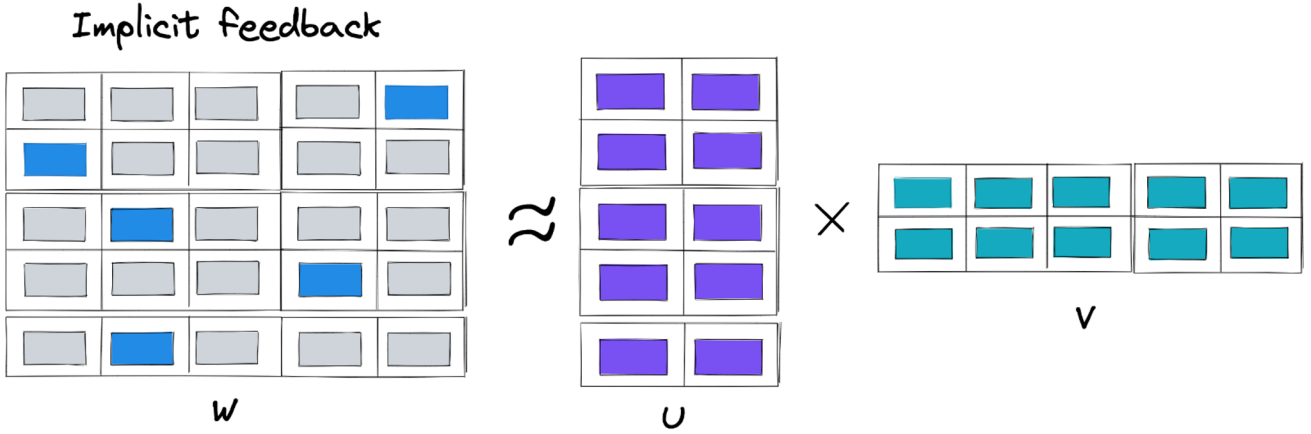
1 Compute exact Pareto front



2 Evolutionary algorithms

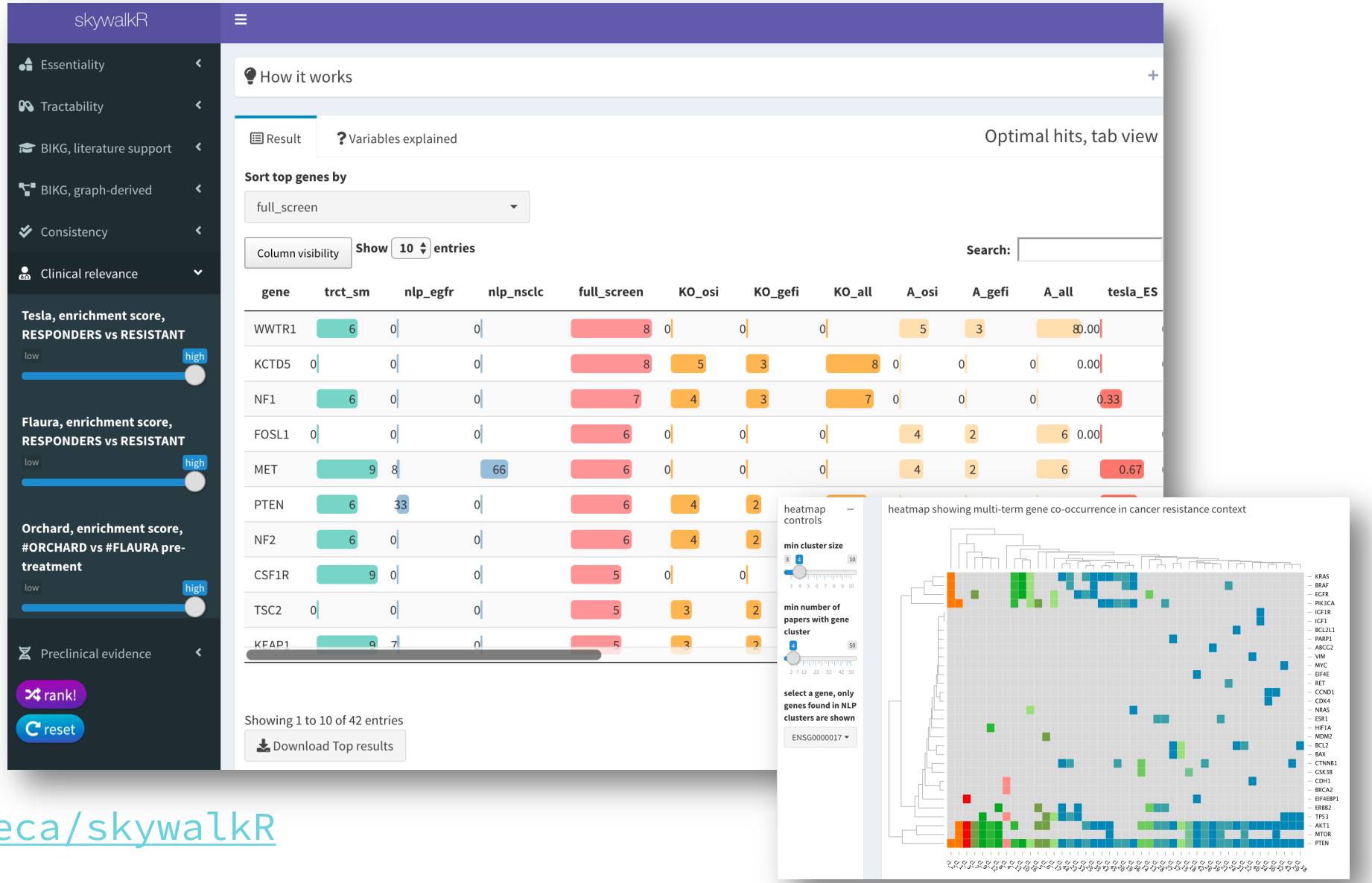


3 Matrix factorization



SkywalkR, interactive interface

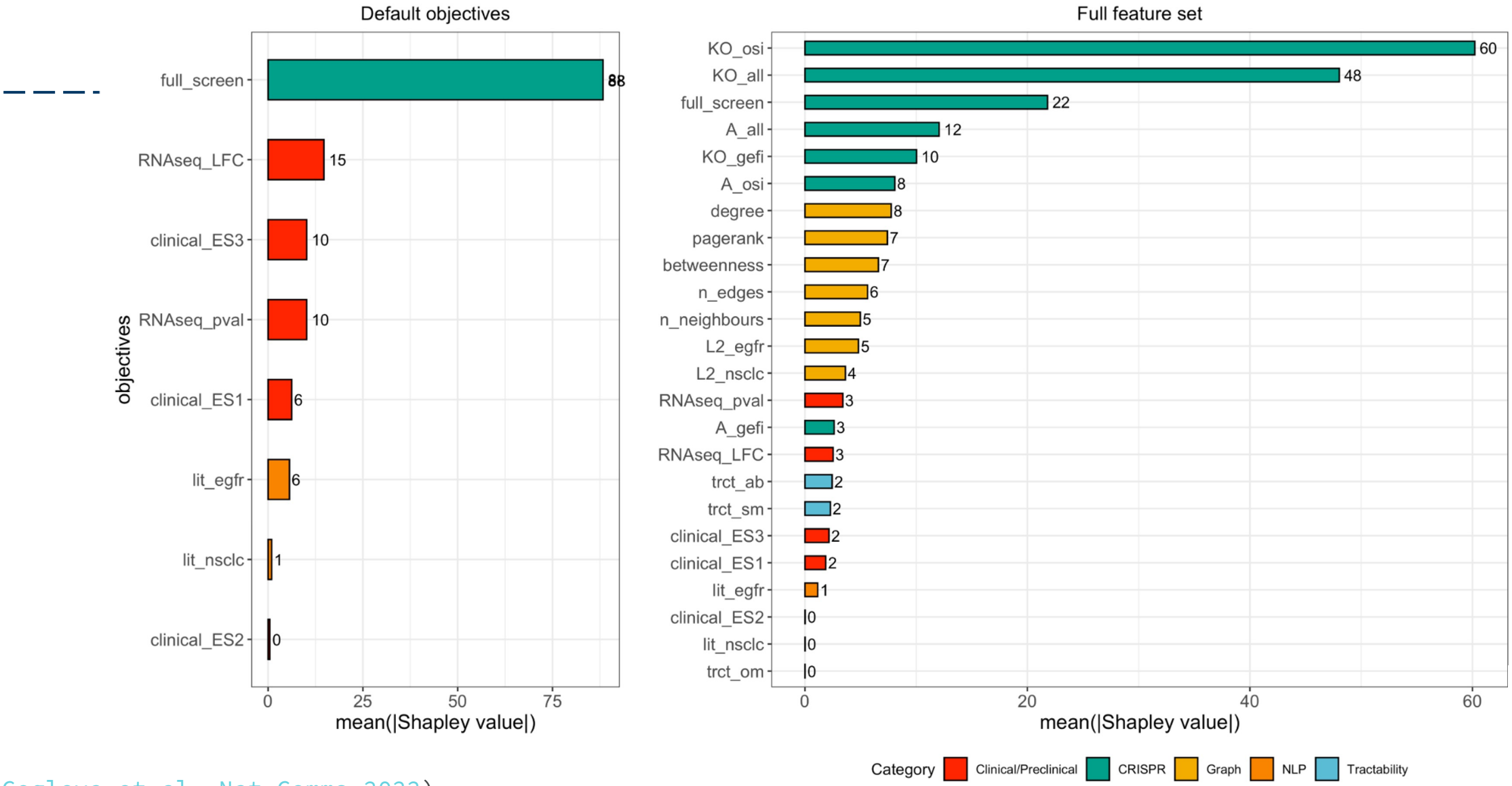
- select a subset of objectives
- set optimization directions
- explore trade-offs



github.com/AstraZeneca/skywalkR

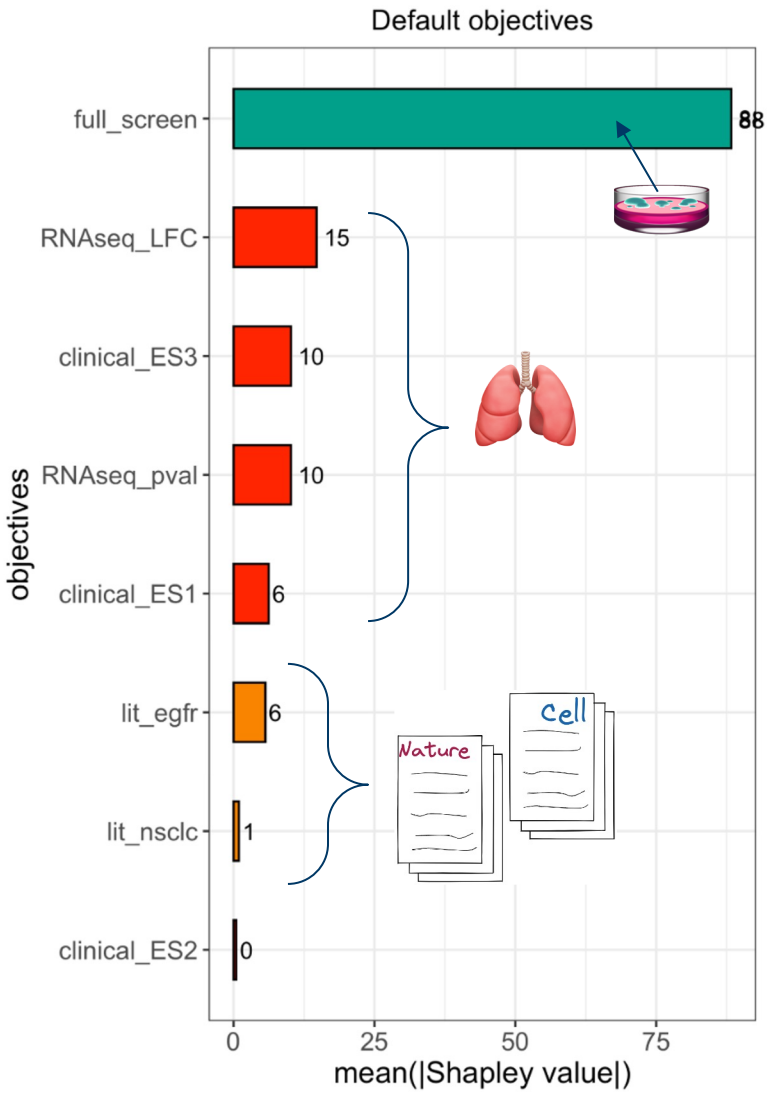
Imperfect validation

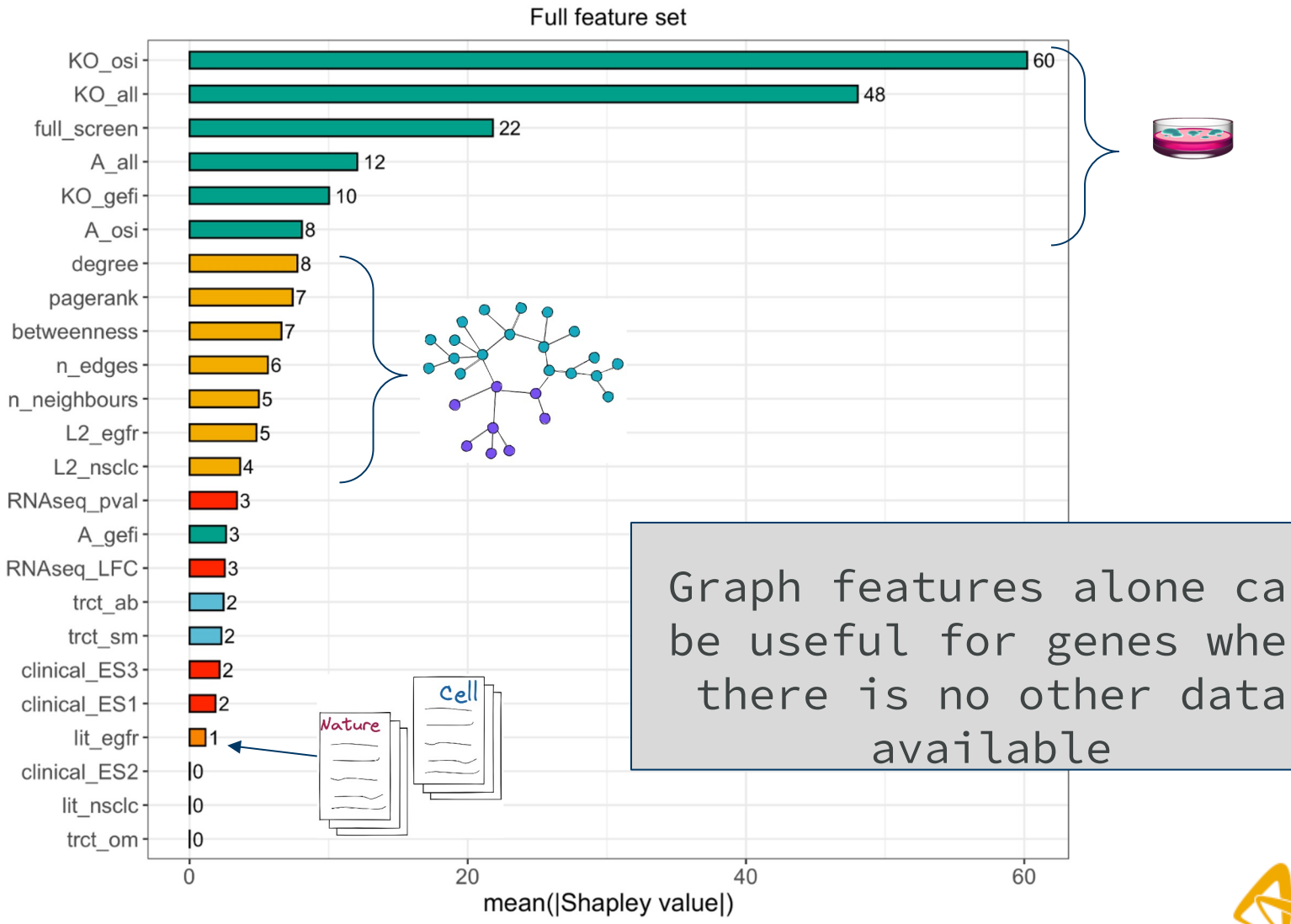
Model domain scientist as a black box classifier



(Gogleva et al, Nat Comms 2022)

Graph-derived features follow clinical in unbiased setting

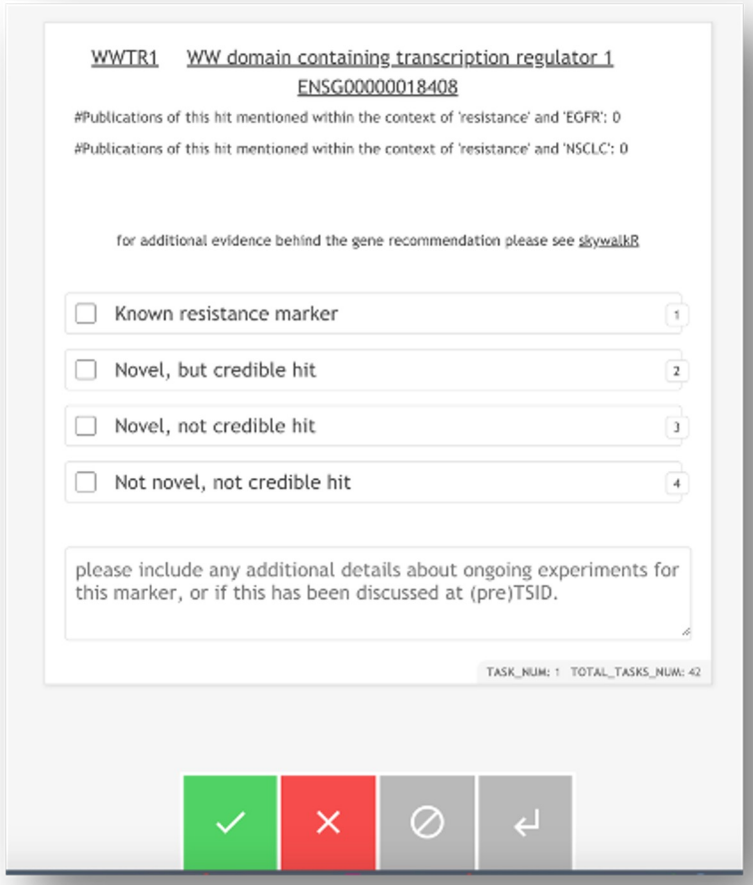




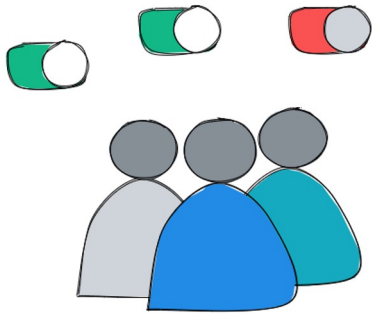
Graph features alone can be useful for genes when there is no other data available

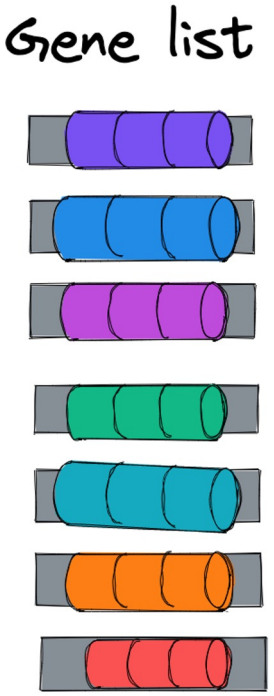
(Gogleva et al, Nat Comms 2022)

Annotation by the experts

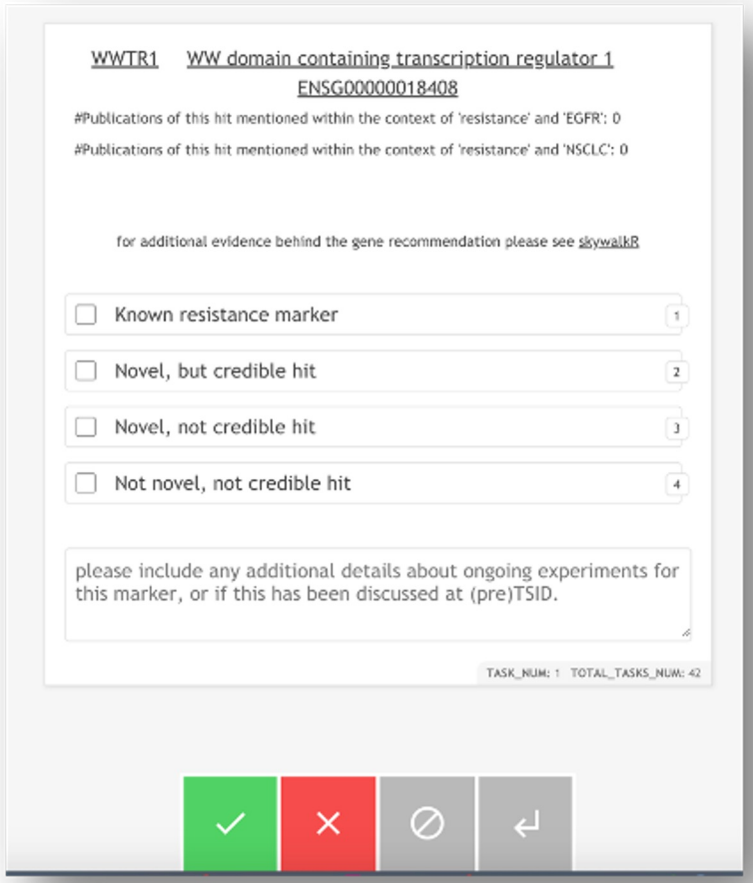


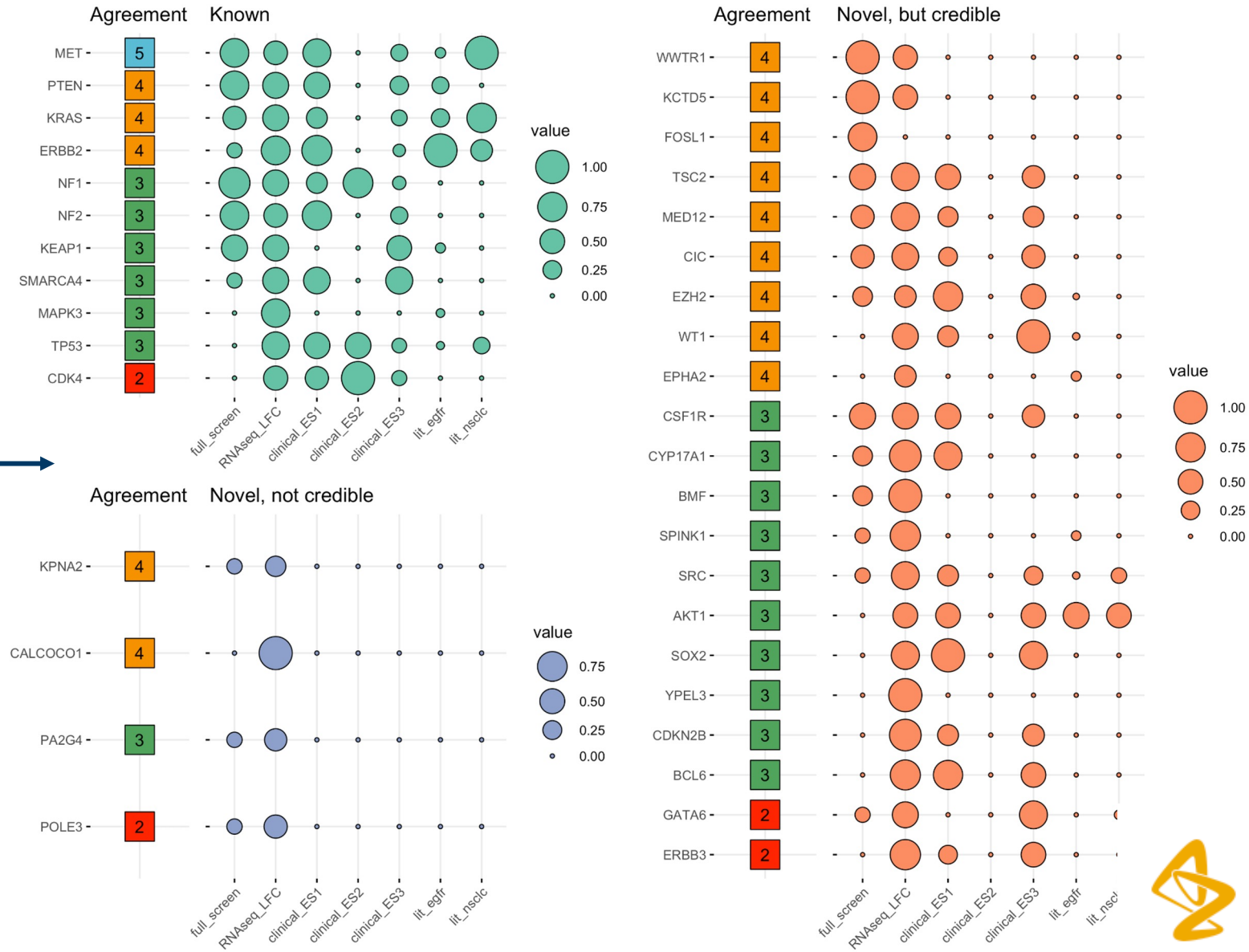
prodigy





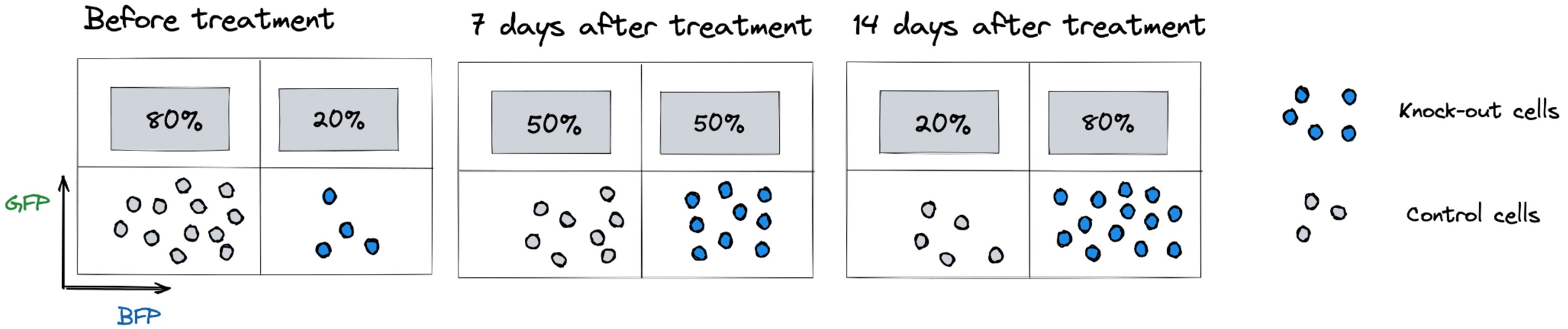
Most of recommendations are ‘novel & credible’



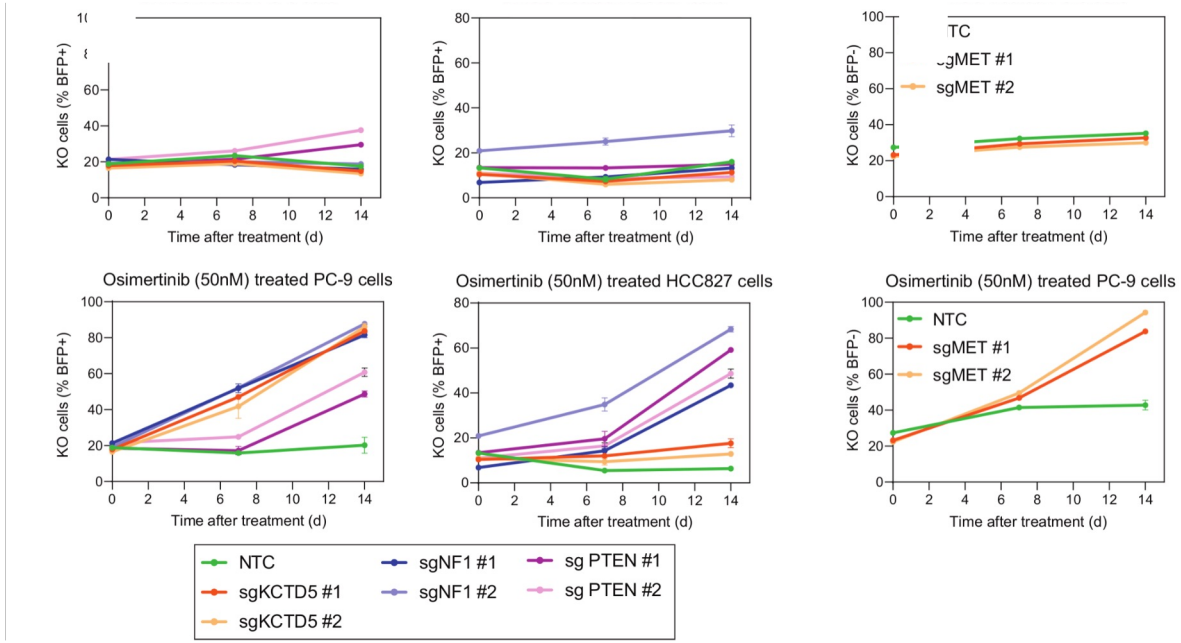


(Gogleva et al, Nat Comms 2022)

Experimental validation *in vitro*

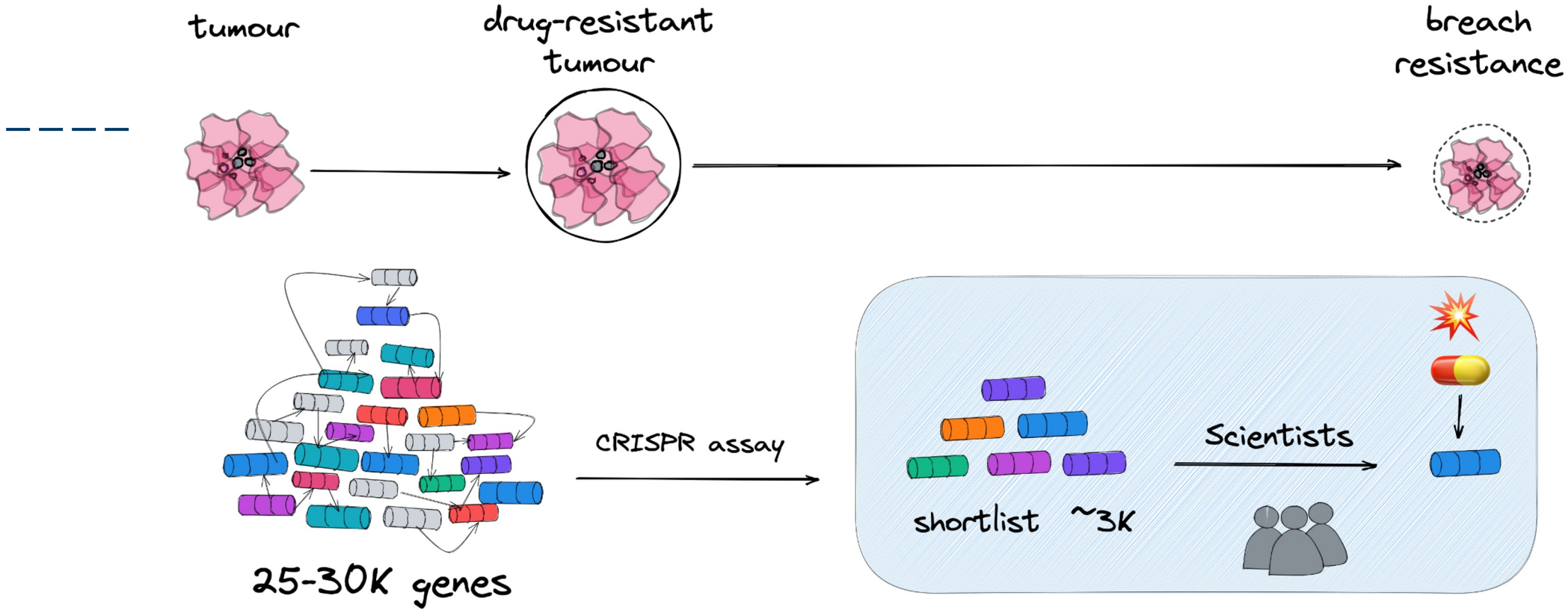


- confirmed involvement of 6 recommended genes in drug resistance
- next: test the remaining genes



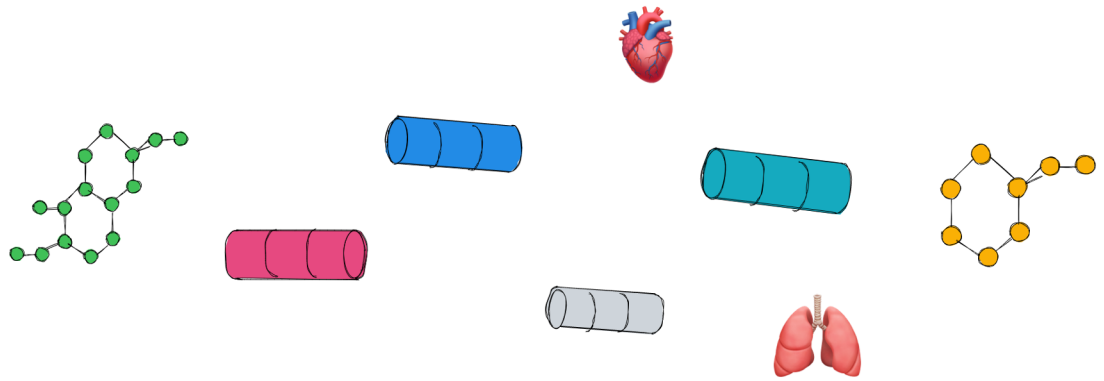
(Gogleva et al, Nat Comms 2022)

Imperfect, yet already useful recommendation system



- 🐌 -> 🏎️ re-rank lists in seconds, not months
- ⚙️ automated feature generation
- ♻️ approach can be re-used in related problems
- 👍 now a standard solution for CRISPR screens

Take home message



- Drug discovery is an exciting field for recommender systems
- Relatively simple recommenders can have a lot of impact
- Need for recommenders that can operate in unsupervised or weakly supervised settings
- There are a number of challenges

Acknowledgements

Early Computational Oncology @ AZ

Krishna Bulusu

Ben Sidders

Daniel Barrell

Miika Ahdesmäki

Jonathan R. Dry

R&D IT @ AZ

Vladimir Poroshin

Michaël Ughetto

Eliseo Papa

Bioscience, Oncology R&D @ AZ

Matthias Pfeifer

Ultan McDermott

We are hiring 🐱💻!